Summer 8-9-2017

The Structural Consequences of Big Data-Driven Education

Elana Zeide

Follow this and additional works at: https://scholarship.law.unc.edu/aidr_collection

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, and the Educational Technology Commons
The Structural Consequences of Big Data-Driven Education

Elana Zeide1–3,*

Abstract
Educators and commenters who evaluate big data-driven learning environments focus on specific questions: whether automated education platforms improve learning outcomes, invade student privacy, and promote equality. This article puts aside separate unresolved—and perhaps unresolvable—issues regarding the concrete effects of specific technologies. It instead examines how big data-driven tools alter the structure of schools’ pedagogical decision-making, and, in doing so, change fundamental aspects of America’s education enterprise. Technological mediation and data-driven decision-making have a particularly significant impact in learning environments because the education process primarily consists of dynamic information exchange. In this overview, I highlight three significant structural shifts that accompany school reliance on data-driven instructional platforms that perform core school functions: teaching, assessment, and credentialing. First, virtual learning environments create information technology infrastructures featuring constant data collection, continuous algorithmic assessment, and possibly infinite record retention. This undermines the traditional intellectual privacy and safety of classrooms. Second, these systems displace pedagogical decision-making from educators serving public interests to private, often for-profit, technology providers. They constrain teachers’ academic autonomy, obscure student evaluation, and reduce parents’ and students’ ability to participate or challenge education decision-making. Third, big data-driven tools define what “counts” as education by mapping the concepts, creating the content, determining the metrics, and setting desired learning outcomes of instruction. These shifts cede important decision-making to private entities without public scrutiny or pedagogical examination. In contrast to the public and heated debates that accompany textbook choices, schools often adopt education technologies ad hoc. Given education’s crucial impact on individual and collective success, educators and policymakers must consider the implications of data-driven education proactively and explicitly.

Keywords: big data; personalized learning; competency-based education; smart tutors; learning analytics; MOOCs

Introduction
We are entering a brave new world of data-driven education. Teachers, administrators, and policymakers increasingly rely upon automated technologies for pedagogical decision-making.1–3 I use “data-driven” to describe systems that operate without human intervention. In this article, I focus on automated instructional technologies that perform core educational functions: delivering instructional material, assessing student progress, and documenting attainment.4 Big data analytics diagnose and predict student progress to inform both instructional and institutional choices. Smart learning environments create new ways to communicate academic attainment. This article considers big data-driven education in public schools, although many of the points mentioned hereunder also apply to publicly-funded private education institutions.

Big Data-Driven Education
As background, I describe different ways schools use data.5 In physical schools, teachers and administrators use student information to inform teaching decisions, give grades, award credits, and create transcripts. Today, faster internet speeds and cloud computing supports similarly interactive virtual learning environments.6 Learners select among modular videos, practice problem sets, and explore supplemental reference material at their convenience. Schools also incorporate these platforms into
their curricula to provide “blended” learning experiences that contain both online and physical components or “flip” instruction, so that students watch lectures as homework and discuss their content in classroom.7

Massive Open Online Courses (MOOCs) are perhaps the most-hyped virtual learning environments. At first, MOOCs offered content from elite universities for free to anyone with an internet connection. They promoted their platforms as a means to democratize high quality education, prompting the New York Times to declare 2012 “The Year of MOOC.”8 Instructional education technologies have evolved since then. The most prominent MOOC providers, Coursera, edX, and Udacity, have all narrowed their focus to vocational and professional subjects and charge users for participation or certification.9 New providers continuously create innovative education experiences through adaptive e-textbooks, mobile apps, and education games.10,11,12

Technology-mediated instruction
Technologically-mediated education technologies generate a continuous stream of information as learners interact with digital platforms.13–16 Students and teachers input information such as usernames, emails, and grade level to set up student accounts. Learners also provide typical academic information through emails, online discussions, assignments, and tests. Digital education technologies collect an unprecedented amount of information about students’ behavior and performance during the learning process—details that could not be recorded or analyzed at scale in physical classrooms. This includes metadata such as time stamps, device identifiers, and even geolocation information.13,17 These systems track the parts of a video students actually watch, when they log in, and how they paused before answering a question.18

The Internet of Things will expand the possible sources for student data exponentially.19,20 Online proctoring platforms use video, facial recognition, audio, and biometric information to verify student identity and detect cheating.21 “Smart” college campuses collect information from radio-frequency identification (RFID) cards that record when students go to the gym and what they bought for lunch.22 Oral Roberts University, for example, requires students to wear Fitbits.23 The most ambitious education reformers and researchers envision a future wherein sensors track students’ eye or breathing patterns to determine their level of engagement or anxiety.24,25

Data-informed decisions
Big data-driven education technologies incorporate and analyze this wealth of information to inform classroom and institutional decision-making.26,27 Learning analytics can provide a more precise diagnosis than harried human instructors.28,29 They track student progress using knowledge maps that break down the relevant subject matter into concepts and “competencies.”30 Platforms, for example, can determine that a student’s poor chemistry grade is the result of failure to grasp a specific algebraic concept the prior year, rather than any difficulty with the scientific concepts themselves.31 These cognitive models may include or infer emotional and cognitive states as well as academic progress.32–33

Most of today’s school technologies interpret and present this information to educators on digital dashboards. The level of interpretation and inference involved in these systems varies widely. Some dashboards show “skill meters” that visually graph learners’ mastery of specific concepts.34,35 Others reduce a complex array of information to sort learners into simple categories.36,37 One early warning tool, for example, uses red, yellow, or green indicators to who teachers students’ likelihood of passing.

Educators can incorporate data-generated student assessments and predictions to support their independent decision making. As Ryan Baker memorably writes, “stupid” tutoring systems can be crafted to inform, rather than replace, “intelligent” human decision-making.38 Many platforms currently in use are primarily oriented at detecting and presenting these patterns to educators. Alt-schools, for example, constantly monitor classrooms to collect digital, audio, and visual information about student interactions and behavior.39 Teachers rely heavily on computer analytics to make sense of this data, still decide what students need next.

Personalized platforms
In contrast to data-informed decision-making, data-driven education systems do not support but supplant human decisions. Instead, computers “personalize” learning automatically by evaluating instructional options in light of students’ profiles and delivering content accordingly.40 In doing so, they try to mimic the way that teachers adapt to student needs in physical settings.

Computers can customize instruction in several ways.32,41,42 Some adaptive systems deliver different material, for example a review of concepts as opposed to practice problems, based on student responses. Others
let students advance through subject matter at their own pace. At Summit schools, for example, students work on a specific concept until they master it, independent of their classmates’ progress. At their most sophisticated, “intelligent tutoring” platforms do more than lead students through predetermined pathways.

As already discussed above, big data analytics often use cognitive models to track and assess student progress. “Smart” learning systems perform a similar process using “tutoring” models that capture different instructional options. The software uses historical data about how students with similar profiles fared to determine the choice most likely to lead to student success. To use a greatly simplified example, say that a student tries three times before entering a correct answer to a practice problem. The system updates a student profile with this information, which creates a data pattern I shall call “ABCD.” The tutoring model includes three different instructional options: it can show a video or have students review the relevant part of earlier lectures. The platform analyzes data created by prior students with ABCD profile patterns which show that 70% of students who watch new videos answer the next set of questions correctly, compared to 55% of those who review old material. The data accordingly “predicts” how the two options will affect the student currently using the system, and, in this case, plays the new video. Smart tutoring systems have yet to migrate into the mainstream, but they are poised to do so.

Visionary Benefits

Smart learning systems promise to promote both equality and efficacy. The U.S. Department of Education and philanthropists including Bill and Melinda Gates and Mark Zuckerberg promote data-driven learning technologies as a means to a more effective, cost-efficient, and equitable education system.

Moving beyond the factory model

Proponents present personalized learning systems as a way to move past the one-size-fits-all factory model of education. Reformers and providers see automated instruction and assessment as a way to improve education quality, particularly in underserved and overcrowded schools. In doing so, they also hope to address disparities in educational achievement and attainment across racial, ethnic, gender, and class categories and create more equitable access to opportunities.

Competency-based credit

Many reformers, including the U.S. Department of Education, want to use data to change how schools measure and document academic success. Algorithmic profiles “embed” assessment seamlessly into instruction instead of periodic, high stakes tests. Schools can award credit and document attainment based on technology-defined competencies, rather than traditional courses, grades, and credit hours. Students can streamline their education acquiring the specific skills they need. Competency-based credentials might also help students from less prestigious schools compete with their peers at more elite institutions.

Independent credentialing

Reformers promote mastery-based assessment and credentials as a means to capture student skills more accurately across institutions. However, competency-based credentials will only be valuable if admissions boards and employers trust in the accreditor. The most ambitious CBE visions seek to employ distributed ledger technology like the block chain to create immutable, self-verifying records. The decentralized nature of these record-keeping systems would also anyone to contribute to students’ credentials, so that learners can accumulate recognized credit for informal learning and life experience outside classrooms.

Efficacy, Privacy, and Equity Concerns

The public conversation surrounding new big data-driven education technologies focuses on their effects, considering efficacy, privacy, and, equity concerns. The following section provides a brief overview of relevant issues as a contrast to the structural shifts discussed below.

Efficacy

Most reformers and schools evaluate big data-driven education technologies in terms of their efficacy in achieving defined learning outcomes. In doing so, they overlook the problematic aspects of implementing big data driven education technologies at scale. The experimental nature of ed tech innovation means it is inevitable that some big data-driven education ventures will fail. The accuracy of outcomes will depend on the representativeness, accuracy, and relevance of the data incorporated into systems as well as the technologies themselves. The size and complexity of data-driven systems makes it difficult to detect errors—and to implement the correct
adjustments upon doing so. How long will it take to discover that a particular tool is not effective or has unequal effect on different populations? Who will be responsible for tracking these and making sure?

Adjusting to flaws and failure will be increasingly complicated, given the highly politicized, bureaucratized, and decentralized structure of the U.S. public education system. It’s one thing to have agile development for software that provides a relatively narrow array of services directly to users, like apps. It is another to have to change fundamental aspects of systems deployed nation-wide and with clients—schools and districts—who may not have the money or resources to keep track of and implement changes. We do not yet have the institutional, ethical, or governance mechanisms set up to grapple with beta education at scale.

Privacy
Increased data collection and school reliance on outside technology providers raise student privacy concerns about access to and commercial use of student information. Big data-driven education systems capture vast amounts of personal and personally identifiable information about students and teachers. Schools share this information with education technology providers who are predominantly private, for-profit entities. This data holds considerable commercial value outside the school context and apart from education purposes. Parents fear that companies will prioritize short-term profits over cautious information use and disclosure.

Traditional student privacy regulation, like the Family Educational Rights and Privacy Act (FERPA) don’t address these issues sufficiently. Although FERPA theoretically provides parents with control over school sharing of personally identifiable student information, its exceptions delegate most data-related decisions to educators. Even in cases wherein parents do have the choice to opt-out of specific classroom technologies, they often do not feel they can do so in practice without putting their children at a significant social and academic disadvantage.

Newer state student privacy regulations try to ensure that vendors use student information appropriately through purpose limitations. Whether regulated directly or indirectly through school disclosure requirements, ed tech providers can only use student information for “school” purposes. As a result, these rules do not impose specific rules regarding school use of big data-driven education tools. Purpose limitations, however, provide minimal protection against problematic aspects of big data-driven education tools used in schools. These laws operate under the assumption that school purposes serve educational interests. They do not account for institutional pressures that may put schools’ interests at odds with students’. Recently, for example, a university president tried to use predictive analytics to determine which freshmen to encourage to drop out in order to improve reported retention rates. In addition, many state laws have an explicit exemption for personalized learning platforms. SOPPIA allows education technology vendors to use covered student information for adaptive or customized services.

Equity
Big data-driven education also offers the promise of more equitable education outcomes, but may inadvertently have the opposite effect in the long run. Data-driven systems promise to be more precise and consistent than humans, which often leads to the presumption that they will be more accurate and objective as a result. Educators, even with the best of intentions, may rely on irrelevant or inappropriate factors in pedagogical decision-making. This can be because of bias toward certain groups or cognitive tendencies that may inadvertently shape decision-making of their conclusions, such as, whether a teacher grades student articles before or after lunch. Machine analysis may offer consistency, but that is not the same as objectivity. Algorithmic analysis can be just as biased as human decisions.

Big data’s predictive tools can similarly help or hinder socioeconomic mobility. Schools can monitor students to identify those at risk of dropping out in time to intervene. At the same time, predictions may be based on historical patterns that reflect existing inequities and discrimination. Predictions can also create self-fulfilling prophesies that unfairly limit future opportunities based on early performance. Long term predictions are particularly problematic in education spaces, which are explicitly environments dedicated to student development. Learning; and life trajectories, are rarely linear. Predictive analytics cannot ‘literally predict [student] life outcomes’ because they cannot incorporate the impact of outside circumstances and student agency.
Structural Shifts
With all the focus on whether a specific technology “works” and who can access student information, educators often overlook the important structural shifts that occur even if technology or policy resolves the above-mentioned efficacy, privacy, and equity issues. These occur even in an impossibly perfect world where big data-driven tools perform as intended, technology providers only use student data to serve school purposes, and data analysis does not reflect hidden bias.

Monitored and memorialized learning environments
Smart learning platforms fundamentally alter learning environments by imposing new information infrastructures. Constant collection, scoring, and memorialization reduces the intellectual privacy characteristic of physical classrooms. The current approach to big data analytics presumes more data are better, leading to the expansion of types of information collected about students and increasingly creating spaces of pervasive surveillance. This constant monitoring has documented chilling effects on student expression, risk-taking, and diversity of opinion. Data-driven education environments accordingly undermine the traditional safety that supports the learning process. Research suggests that students’ sense of vulnerability impedes academic promise and disproportionately affects minorities.

As already discussed above, data-driven learning platforms continuously assess student progress. While teachers do the same in physical classrooms, technologically-mediated education environments collect and capture of students’ experiments and mistakes during the learning process. This collapse of formative feedback, summative assessment, and credentialing raises the stakes of every mistake or misstep. Students’ every action might be incorporated into the digital equivalent of transcripts. Students’ early mistakes can be preserved for later scrutiny and mined for new algorithmic inferences.

The prospect of preserving these records using block chain technology ratchets up the stakes even more. The open nature of the block chain makes these truly public permanent records. This runs counter to the consistent theme in U.S. society, economic policy, and political rhetoric that past should not unduly limit future opportunities. Like the expungement of juvenile criminal records or old bankruptcy proceedings, the practical obscurity of classroom proceedings promotes what Andrew Tutt refers to as “revisability.” The surveil-

Displaced pedagogical decision-making
Schools outsourcing pedagogical functions to companies outsource important decisions to them as well. The networked flow of information replaces the transparency, autonomy, and accountability expected in education spaces with standardized and decontextualized decision-making. Changes in educational evaluation and credentialing shifts the power dynamic to the entities creating competency models and evaluative systems. In performing these seemingly mundane processes, technologies—and their corporate providers—in fact exercise significant authority over fundamental aspects of education that have previously been invested in teachers, school administrators, and (often local) policy makers.

Automated education tools also reduce the autonomy of on-site educators. While personalized learning tools may be more “customized,” they still create standardized systems. Because algorithmic analysis relies on probabilities, data-driven instruction, evaluation, and credentials cannot reflect or react to the unique aspects of specific circumstances. Resulting path dependencies limit educators’ ability to deviate based on highly contextual circumstances to the degree that educators and institutions defer to algorithmic determinations. This cuts against the highly contextualized decision-making characteristic of physical classroom settings. It goes against the idealized education values espoused by big data-oriented reformers as well as their critics, who seek to treat students equally regardless of their group affiliations.

By relocating the site of pedagogical functions, data-driven education technologies make it more difficult for students, parents, and communities to exercise agency and demand accountability. Instead of readily available teachers and administrators, stakeholder must navigate remote corporate communication structures. Those who do obtain information about decision-making may not be able to make sense of the complex algorithms and probabilistic decisions driving personalized learning. The lack of transparency and obvious sources to exercise agency or ensure accountability may exacerbate the alienation of students and parents who already feel disconnected and disempowered in the traditional system.
Inconsistent implementation of big data-driven education technologies creates the risk of a two-tier system. Underserved schools may lack the resources to provide more flexible instruction and assessment compared with better funded counterparts, and rely almost exclusively on the determinations of smart education technologies. Students enrolled in schools which can afford to allow teachers to deviate from algorithmic recommendations or supplement automated assessment with personal evaluation will have the opportunity for accommodation based on individual circumstances. Less fortunate students may receive automatically differentiated instruction without the flexibility of contextualized assessment.

Computable competencies and priorities

Digital mediation of student— instructor communication changes more than the mode of delivery. It changes the content, metrics, and goals of education itself. This makes the stakes much higher than, say, the information practices governing a basic commercial transaction. In virtual retail environments, for example, information infrastructures that shape customers’ shopping experiences and perhaps their choice of widget, but do not alter the nature of the widget itself. In learning environments, information technologies perform crucial functions, and, accordingly, shape education’s content, metrics, and values.

Competency maps and measurement systems are not the result of a transparent reflection of reality. Traditional grading involves human interpretation—whether through informal mental processes or through explicit rubrics. Automated commensuration of students’ educational experience is similarly interpretive. The categorization of tasks and knowledge, the criteria for reaching “competency,” and the metrics used to track and measure students’ achievement all involve value-laden decisions about relevant information, learning processes, and desired outcomes.41,82,107,108,136,142 These choices end up creating the epistemology that then defines education.

The hardware and software used to collect, capture, analyze, interpret, and store student data also limits the content measures, and format of the student scores and credentials.108,143 These tools only measure and, accordingly, can only respond to the features or variables factored into the algorithmic process. This gives short shrift to the psychosocial skills research that increasingly shows to be essential for education attainment.125,133,134 Big data-driven education platforms focus on student mastery of skill and knowledge acquisition. In doing so, they implicitly define education as a collection of expertise and demonstrable abilities.89,107 Although some technologies attempt to measure meta-cognitive abilities, computer platforms cannot currently capture the “soft” skills, like teamwork, crucial for long-term success.125,133

Data-driven education systems also have the potential to inadvertently shift education away from paradigms that promote unquantifiable values to solely instrumental ones.144–148 Schools in America have historically served a plurality of purposes, including cultivating civic participation, promoting socioeconomic mobility, and encouraging intellectual fulfillment.4,149

Overreliance on big data discrimination education technologies risks reducing these more abstract goals to an afterthought.

Conclusion

As pervasive data collection and mining to feed learning analytics creates ubiquitous surveillance, these consequences will impact more and more of everyday life outside of formal education. They change the data used, the evaluation mechanisms, and the ultimate format or records used to assess and represent student achievement, academic credit, and intellectual mastery. The value judgments and commensuration inherent in these systems are often opaque and may be inadvertent, but have important consequences on what “counts” as and toward education and achievement in academic and employment environments.

Each shift in pedagogical decision-making has the potential for unintended consequences because of inaccurate or unrepresentative data, algorithmic bias or disparate impact, scientism replacing more holistic and contextualized personal evaluation, and the exclusion of noncomputable variables and nonquantifiable learning outcomes. Examining big data-driven education in light of structural dynamics teases out the agendas that are being advanced—intentionally or otherwise—when adopting data-driven education technologies.

It is important that the changes wrought by big data in education are not made unknowingly and inadvertently by thoughtlessly implementing new technologies. They should instead be the result of considered choices sufficiently transparent to permit public scrutiny. This may mean requiring more transparency, accountability, or precautionary approaches to information and privacy practices in learning environments. Just as we adopt new approaches based on technological
affordances, schools must also implement accompanying oversight and governance structures that match them.

Author Disclosure Statement
The author received support from Microsoft for research related to privacy in early massive open online courses as part of her prior affiliation with New York University’s Information Law Institute.

References


64. Vian, K. Own Your Achievements: Three Ways Blockchain Tech is Disrupting Education. Campus Technology (5/16/16). at <http://campustechnology.com/articles/2016/05/16/how-blockchain-will-disrupt-the-higher-education-readiness-gap>


96. Selwyn, N. Distrusting educational technology: critical questions for changing times. (Routledge, Taylor & Francis Group, 2014).
105. Williamson, B. in Surveillance Futures: Social and Ethical Implications of New Technologies for Children and Young People (2016). at <https://books.google.com/books?hl=en&lr=&id=S8LDAAAAQBAJ&pg=PT64#v=onepage&q=Ti5z%0An788%0Ah%0A&f=false>
122. Selwyn, N. Distrusting educational technology: critical questions for changing times. (Routledge, Taylor & Francis Group, 2014).


Abbreviations Used
CBE = competency-based education
FERPA = Family Educational Rights and Privacy Act
MOOCs = massive open online courses